DOCUMENT RESUME

ED 246 998 PS 014 396

TITLE	Income Verification Pilot Project (IVPP): The Development of an Error-Prone Model for the School Meal Programs. Revised.
INSTITUTION	Applied Management Sciences, Inc., Silver Spring, Md.
SPONS AGENCY	Food and Nutrition Service (DOA), Washington, DC. Office of Analysis and Evaluation.
PUB DATE	Aug 83
CONTRACT	53-3198-153
NOTE	56p.; For related documents, see PS 014 394-397.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Elementary Secondary Education; *Eligibility; *Error Patterns; Federal Programs; *Food Service; Models; School Activities; *Statistical Analysis
IDENTIFIERS	*Error Detection; Income Verification; Quality Assurance; *School Nutrition Programs

ABSTRACT

This report describes efforts made in 1931-82 to develop an error-prone model (EPM) to help judge the extent of misreporting of income and family size on applications for government-sponsored school meal benefits. (EPM's are statistical formulas that produce scoring systems used to distinguish applications likely to result in excess benefits from applications likely to be accurate.) A sample of 2,480 cases from the National Evaluation of the School Nutrition Programs (NESNP) database was used in the analysis: 1,984 cases were used to develop the model, and 486 were set aside to test the model after its development. The statistical procedure used to generate the initial EPM, sequential search, made use of the Automatic Interaction Detector program. Although analysis of the NESNP data successfully produced an EPM, the model had two major deficiencies. Therefore, data from the 1981-82 Income Verification Pilot Project (IVPP) were used to continue exploring EPM development. The random sample consisted of 741 families from 54 schools participating in Phase I of the IVPP. Results showed that the model developed from the NESNP database was validated on in-home audit data. The goal of producing a simple but effective EPM also was achieved. The resulting scoring profile was found to be capable of selecting a subgroup of applicants who have a 70 percent probability of receiving excess benefits. (RH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☒ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED246998

INCOME VERIFICATION PILOT PROJECT (IVPP)

THE DEVELOPMENT OF AN ERROR-PRONE MODEL FOR THE SCHOOL MEAL PROGRAMS

Revised August, 1983

Submitted to:

Office of Analysis and Evaluation
Food and Nutrition Service
U.S. Department of Agriculture

Submitted by:

Applied Management Sciences, Inc.

This report was prepared pursuant to Contract No. 53-3198-153 with the Office of Analysis and Evaluation, Food and Nutrition Services, U.S. Department of Agriculture. However, the content does not necessarily reflect the position of that agency, and no official endorsement of these materials should be inferred.

PS 014396

EXECUTIVE SUMMARY

This report presents efforts made in 1981-82 to develop an error-prone model for USDA's school meal benefit programs. A 1980 study by USDA's Office of the Inspector General and in-home audits conducted as a part of the Income Verification Pilot Project revealed that misreporting on meal benefit applications resulted in the award of excess benefits to a substantial number of students. Therefore, it was appropriate for the Income Verification Pilot Project, a Congressionally-mandated study of procedures to prevent and detect misreporting, to consider the development of an error prone-model.

Error-prone models (EPMs) are statistical formulas that produce scoring systems, which in turn are used to distinguish applications likely to result in excess benefits from those that are likely to be accurate. Many other Federal benefit programs use error-prone models to reduce fraud and waste. The school meal environment, however, has several features that prevent an easy adaptation of the EPM systems used by other Federal agencies:

- Applications are processed during a brief period at the beginning of the school year. Therefore a complex multi-step scoring system would not be feasible.
- Most school food authorities do not use computers to process applications. Therefore, an EPM would have to be simple and easy to use.
- Maximum meal benefits per applicant equate to a few hundred dollars. Therefore, the model should be highly efficient in identifying applications likely to result in excess benefits. Further, the model should require minimal resources for implementation.

- The school meal application contains a limited set of variables, and school staff are not likely to have the resources to access other student records when reviewing the application. Therefore, the EPM scoring system should rely exclusively on the application.

Given these environmental constraints, the 1981-1982 efforts to develop an error-prone model were viewed as exploratory. The analysis leading to model development proceeded in an iterative fashion. The first analysis objective was to determine whether meal application-type data would be sufficient for producing a model. If the first goal was achieved, the second goal would be to build a model that would be highly effective, yet simple enough to be feasible in the school meal operating environment.

The results obtained in response to these analytic objectives are:

- It is feasible to develop a preliminary error-prone model that uses application-type data and is effective in identifying applications likely to result in excess benefits.

- The analysis leading to the model specification suggested two major changes to improve the meal application process:

  -- Applicants should be required to report income by source for each adult. Requesting aggregated income by household or individual tended to produce omissions.

  -- Requiring receipt of Food Stamp benefits on the application is ideal for distinguishing truly eligible applicants from those who misreport. Receipt of Food Stamps was the strongest predictor of accurate reporting.

- The analysis leading to model development also revealed that omitted or underreported wage income was the largest factor contributing to the receipt of excess benefits.

- The error-prone model developed and refined in 1981-1982 produced a scoring system that is highly effective, yet simple to use. The key predictors of excess benefits in the four-step scoring system are:

  - receipt of Food Stamps, and
  - reporting total income close to the eligibility cut-off point.

- If the error-prone model contained in this report is used to select a three percent sample of applicants, cases with excess benefits are four times more likely to be identified than would be the case using random sampling.

These findings collectively suggest that an error-prone model is an effective and feasible tool for preventing and detecting the award of excess meal benefits. In 1982-83 the EPM will be further tested and refined with a more comprehensive data base.

TABLE OF CONTENTS

Chapter		Page
	Executive Summary	i
1	Introduction	1
2	Development of A Preliminary Error-Prone Model	10
3	Refine of the Error-Prone Profile Based on 1981-82 IVPP Data	28
	APPENDIX A: PREDICTORS USED IN THE ANALYSIS	
	APPENDIX B: ANALYTIC PARAMETERS	
	APPENDIX C: OPERATIONAL DEFINITIONS	

LIST OF EXHIBITS

Exhibit		Page
1	Error Rates in Award of Benefits For School Nutrition Programs	16
2	Variables Related to Misreporting	18
3	Error-Prone Model Tree Diagram	21
4	Groups With Above Average Proportion of Errors	23
5	Groups With Below Average Proportion of Errors	25
6	Validation of NESNP-Defined Error Prone Groups Using NESNP Data and IVPP Audit Data	31
7	List of Variables Used in Development of Application-Based Error-Prone Profile	33
8	Percentage of Households Receiving Excess Benefits By Difference of Reported Monthly Income From Eligibility Threshold	34
9	Application-Based Error-Prone Profile Scoring System	36
10	Application-Based Error-Prone Profile	37

1

INTRODUCTION

This report presents the results of efforts to produce a mechanism that school districts might use to select meal benefit applicants for verification in an efficacious manner. "Error-prone models" refer to statistical formulas that use application characteristics or other accessible data to estimate the likelihood that excess benefits are (or would be) received by an applicant. A major area of investigation in the 1981-82 phase of the Income Verification Pilot Project (IVPP) was the feasibility of developing a preliminary error-prone model, and, if feasible, the development of such a model for testing in 1982-83. The remainder of this chapter presents background information concerning error-prone models and their feasibility in the school meal operating environment. Succeeding chapters present the 1981-82 efforts in error-prone model development and their results.

An Introduction to Error-Prone Modelling and Implementation by Federal Agencies

As was mentioned above, error-prone models are statistical formulas used to identify characteristics or features of benefit applicants and/or recipients that are associated with misreporting. Error-prone models are described non-mathematically as error-prone profiles, and the profiles may be used to screen and select applications for verification.[1] Thus, error prone models and profiles are primarily a detection method because they help to identify a case with errors rather than prevent the applicant from making an error. The development of error-prone models (EPM) however, often provides insights into the nature of misreporting. This information can then be used by

program managers to change the application document, the application process, and/or verification procedures so that errors can be prevented. The EPM development efforts in IVPP produced both types of results.

The use of an error-prone model offers several advantages. First, it provides an objective basis for selecting applications for verification. In contrast to experiential or intuitive selection based on suspicions, an error-prone model offers a non-biased approach to the identification of applications that warrant further examination. Second, selecting only the cases that are most likely to contain errors allows program managers to greatly increase the effectiveness of verification activities. In contrast to random sampling of applicants, an error-prone model permits limited resources to be applied to those applications where the pay-off in errors found and savings achieved are likely to be greatest. Further, error-prone model-based selection excludes from verification those applicants who have very low probabilities of misreporting, thus reducing applicant burden and preventing wasted resources.

Many Federal agencies currently use error-prone models as a tool to support the proper management of their programs. Among them are the Department of Agriculture's Food Stamp Program, the Department of Education's Pell Grant (student grants for higher education) Program, the Social Security Administration's Supplemental Security Income Program, and the Internal Revenue Service's Federal Income Tax Program. A brief synopsis of the role of EPMs in these other Federal agencies follows.

FNS has funded research projects in error-prone modeling for the Food Stamp Program. In North Carolina, an error-prone model is used to provide a basis for targeting the work of personnel involved in verification of applications. In New Mexico, an error-prone model has been developed to identify the most significant types of errors and their causes. New Mexico is now examining cost-effective methods for correcting these types of errors. An error-prone model developed by Texas is being used to identify low-risk applicants who can be given extended certification, thereby permitting reallocation of administrative resources to the applications that are more likely to contain errors.

The Social Security Administration uses an error-prone model in managing the Supplemental Security Income (SSI) Program. The model specifies the average dollar error for different types of beneficiaries and the results are used to establish redetermination procedures for each beneficiary, based on the expected dollar loss. In cases where large losses are anticipated because of the high probability of error, redeterminations are conducted more frequently and involve more detailed processing. The model is designed to direct redetermination efforts to the types of benefit cases where the expense of processing is most likely to be offset by savings resulting from the detection of relatively costly errors.

The U.S. Department of Education, Office of Student Financial Assistance uses an error-prone model in managing its Pell Grants Program. This model is used to prevent the misallocation of Federal student financial aid funds. While processing the applications, those that match a profile for a high probability of misreporting are sampled for verification. These applicants are asked to provide documents to support information shown on their application forms. Selection of applications is performed by computer after eligibility is determined and verification is performed by local college financial aid administrators. No benefits are awarded until application data are verified or corrected.

In contrast to the pre-award review performed by the Office of Student Financial Assistance, the Internal Revenue Service uses error-prone modeling to review tax returns after initial processing. Once returns have been processed and refunds have been made, IRS selects for audit tax returns matching a profile suggestive of a high probability of error. Taxpayers are then contacted and undergo an audit. If errors are found, the taxpayer makes payment, plus interest to the IRS. Thus, although IRS performs its audits after return processing, the penalty and payment requirements equate their system to a pre-award detection system.

As is evident from this brief discussion, Federal agencies use error-prone models for a variety of purposes:

- To gain a better understanding of error sources so that corrective actions can be taken to prevent errors.
- To target administrative resources so that savings for error detection surpass the expenses of performing verification.
- To reduce applicant burden by eliminating from verification efforts applications and/or beneficiaries likely not to be in error.

These examples also represent a variety of operational environments. For instance, the Pell Grant Program is administered centrally by the Federal government. Students apply directly to the government and applications are processed by a contractor in a central and satellite sites. EPMs are applied by computer as a part of application processing, and then local college financial aid administrators conduct verification before awards are disbursed. In other programs, the state plays a role or the application process occurs at the local level. For instance, in programs such as AFDC and Food Stamps, where the state contributes to benefits and administrative costs, state agencies often are responsible for developing and updating the EPM, and overseeing its use by local agencies. Profiles are provided to local offices where caseworkers manually screen applications or, if the application process is computerized, the applications are screened by computer. In programs such as Food Stamps, AFDC and SSI, all applications are verified[2] before benefits are awarded, and error-prone models are used to:

- Identify cases for more frequent or less frequent redeterminations.
- Identify cases likely to require more extensive or less extensive verification.

None of the programs highlighted here operate exactly as the school meal programs do. The school meal programs are similar to Pell Grants in that the local administrative staff are not paid directly for helping the Federal government administer the program. This occurs because the programs themselves are thought by some to produce an income for the educational institution. In the case of Pell Grants, the program makes higher education affordable to students who otherwise would not be able to attend, thus supplying more income from tuition fees, etc. This

income is used to acquire and maintain facilities, and to provide educational programs and services. In the case of the school meal programs, the Federal reimbursement for meals and student payments for meals provide a significant level of funds needed to support the program's administration and operation. In direct contrast, public assistance programs like AFDC, SSI, and Food Stamps do not produce an income for the local administering agency. Instead, the role of the local agency is to process applications and disburse monetary or equivalent benefits.

The Pell Grant program is also similar to the meal programs in that the majority of benefit determinations occur in the beginning of the school year. In contrast, other Federal benefit programs normally process applications at a relatively uniform rate throughout the year.

Although most similar to the Pell Grant Program, the school meal programs differ from Pell Grants in one important respect. In the school meal programs application processing and eligibility determinations are performed by the local agency that is also responsible for disbursing the meal benefits. In the Pell Grant Program, local agencies are responsible only for disbursements; the Federal government is responsible for reviewing and approving the application. In this respect the school meal programs are more similar to AFDC and Food Stamps because local staff handle application processing, eligibility determinations, and disbursements.

These similarities and differences demonstrate that no other Federal program provides a precise example of EPM application that could be transferred to the school meal programs. A summary of issues associated with adapting any EPM system for the school meal environment follows.

Applying Error-Prone Models in the School Meal Operating Enviroment

The use of EPMs in the school meal programs is being considered by IVPP because there is a substantial amount of error in the meal application process and as a consequence public funds are being used to provide benefits that exceed the true eligibility of some applicants.

The error rate is documented in earlier IVPP reports[3] and in previous research by USDA's Office of the Inspector General. Therefore, it is incumbent upon program administrators to take corrective action, and it is advisable that the corrective action be as cost efficient as possible. An error-prone model strategy is a likely candidate to support the proper management of the programs. However, the school meal programs have several unique features that affect the feasibility of implementation:

- Families are required to reapply for meal benefits annually. Because schools traditionally complete administrative paperwork at the beginning of the school year, virtually all meal applications are processed before October of each school year. This places a burden on determining officials for quick review and processing. In many other Federal assistance programs eligibility determination and verification occur throughout the year. Therefore, the burden of conducting verifications based on an error-prone model is spread across the year rather than occurring all at one time. This concentrated workload problem is unique to the school meal programs and the Pell Grant Program.
- Traditionally, the school meal programs have accepted self-reported, unverified information about household size and income and have based eligibility determination on these reports. In contrast, many other Federal assistance programs require documentation of income to support the eligibility determination process. The imposition of an error-prone model and additional verification requirements may meet resistance from both parents and SFA officials.
- The enabling legislation mandates that the application form for participation in the school meal programs must be easy to complete. The amount of information that can be requested is, therefore, limited and the form must be easy to understand. The need to use an uncomplicated application affects flexibility to add screening items that would support identification of potential misreporters.
- The school meal programs have a relatively low per student dollar benefit rate--approximately $200 per student annually. Efforts to reduce misreporting will not be cost-effective unless the amount saved through prevention and detection of misreporting is great enough to offset the additional administrative costs associated with verification. Cost-effectiveness may be more readily achieved in other Federal assistance programs where benefit levels are higher.
- Applications for meal benefits are processed by SFA and/or other school officials who have many other responsibilities and

limited time to devote to the eligibility determination process. Implementing an error-prone model would place an additional burden on these personnel. In contrast, some other Federal assistance programs assign personnel to application processing, eligibility determination, and verification as a major aspect of their job. This feature has implications for the cost and feasibility of implementing additional verification requirements based on the results of an error-prone model.

- School food authorities do not receive any Federal or state funding to conduct eligibility determination and verification efforts. An error-prone model may introduce additional staff requirements that are not currently funded. SFA officials believe that they may not be able to perform verification unless they receive additional resources.
- There is no provision in the law that would permit recovering funds from individuals who received free or reduced-price meals for which they were not eligible. Other programs have such provisions or are able to adjust benefits to reflect the results of verification. Consequently, error-prone models will have the greatest pay-off if they are used early in the school year before substantial benefits have been awarded.

These features create some major challenges not only for the development of error-prone models, but also for virtually any form of corrective action. Given these circumstances, it was decided that the investigation of error-prone models in IVPP would be most sensitive to the following issues:

- Because the maximum amount of benefits that could be overawarded is so small, the model must be able to target in on highly error-prone groups that would make verification cost effective.
- Because of limited SFA resources and the lack of widespread computerized processing systems, the model should be quick and easy to implement in a manual processing environment.
- Because legislation mandates a simple application, the model should ideally rely on existing application data.

These issues dictated an iterative process for model development that responded to the following questions.

- First, could an effective model actually be developed with the limited data currently available and potentially available on the application? That is, could a model be developed at all, and if so, would it be effective in identifying groups of applicants with an extremely high or extremely low error

rate?

- Second, if it was possible to develop a model, could one be both effective in detecting applications with high error rates and be non-burdensome to use?

These questions are addressed in Chapters 2 and 3.

END NOTES

[1]For the purpose of simplicity, the terms error-prone models or EPMs, refer to both models and profiles.

[2]Pre-award verification methods used by other agencies include: requiring documentation with the application; collateral contact with employers, banks and other local agencies; and inter-county and intra-state tape matches or data exchanges with employment security wage files, motor vehicle files, child support and enforcement records, etc.

[3]The IVPP School Year 1981-82 In-Home Audit findings (Applied Management Sciences, May 1983) found that about 17 percent of the applicants for school meals had application errors that produced an erroneous eligibility determination. A 28 percent error rate was found by USDA's Office of the Inspector General in 1980.

2

DEVELOPMENT OF A PRELIMINARY ERROR-PRONE MODEL

At the beginning of the IVPP in 1981, FNS possessed two nationally representative data bases that potentially could be used to attempt to develop an error-prone model. Neither data base was ideal, however, because neither contained both verified eligibility data and application data.[1] The two available data bases were:

- An OIG audit data base containing information about 766 meal program participants collected in May of 1980.
- The National Evaluation of the School Nutrition Programs (NESNP) data base containing information about 7,628 meal program participants collected in the Spring of 1981.

The primary advantage of the OIG data base was that it possessed eligibility errors detected by audits. For the purpose of EPM development, however, it was limited because it was a relatively small data base and contained few variables from the application that could be used as predictors. For instance, it did not contain children's grade or family income sources. Further, the data were older than the NESNP data, and did not reflect program changes that occurred between 1980 and 1981. This fact would reduce the generalizability of the EPM results.

The NESNP data base also had a major limitation relative to EPM development. The criterion measure (award of excess benefits) was based on personal interviews, not audits. However, the recency of the data, the large data file, and the availability of numerous variables to be tested as predictors led to the selection of NESNP for the development of a preliminary EPM.

Because NESNP was a comprehensive evaluation of the school meal programs, all data available from NESNP were not relevant to EPM development. Therefore, a sample of 2,480 cases that met the following conditions were considered for analysis:

- The child was reported by the parents as receiving free or reduced-price breakfasts or lunches. (NESNP also included children who were ineligible and children who received milk only.)
- Adequate data were available to permit an eligibility determination to be made. (Since NESNP was based on interviews with parents, the application was not available and eligibility had to be computed from the interview data.)
- The computed eligibility either matched or exceeded the parent's report of eligibility. (Since application-based eligibility was not available, eligibility was computed from the income and household information reported during the interview. Only those cases where the reported eligibility matched or exceeded computed eligibility were used in analysis. We were not interested in cases where the computed eligibility was less than reported eligibility because we were primarily interested in identifying cases of excess benefits, not insufficient benefits.)

Of the sample of 2,480 cases, 1,984 cases were used to develop the model and 486 cases were set aside to test the model after it was developed. Thus results based on the sample of 1,948 should replicate in the sample of 496 if the model is generalizable.

The Predictive Variables Selected for Testing

There are two reasons why a variable may efficiently predict misreporting and excess benefits. First, the variable may be related to characteristics of the individual that may make him or her careless or dishonest or that may make it difficult for the applicant to determine his or her own income. Such a characteristic could be self- employment, for example, because self-employed individuals are likely to have irregular and unpredictable incomes. Second, the variable may be related to a person or family having an income above the income eligibility cut-off. Owning a home might be such an indicator.

Predictors were selected from among the many present in the NESNP data base if they met one of the following criteria:

- The predictor could be easily accessed from the application or school records and could, therefore, be used to screen applications on the basis of an EPM.
- An applicant would be unlikely to distort the information

in the application, making it a reliable item to use in screening applications.

An example of a predictor that meets the first criterion would be number of siblings in school, which could be obtained from the application (in some instances) or from school records. A second example is location of home (urban, suburban, small town or rural), which could be determined from the address or location of the school. Parents' occupation (or at least whether they had a non-farm labor occupation) could probably be determined from other records independent of the application (e.g., parent emergency contact information maintained by the schools). Clearly, some of these variables would be easier to access than others.

The second criterion (i.e., variables likely to be reported accurately because their connection to eligibility determination is not immediately obvious) is somewhat less precise. When asked, most people will describe their occupation accurately, though they might not if asked in a context where they thought it could affect their eligibility. The same may be true of home ownership. Variables such as participation in the Food Stamp Program or other assistance programs are unlikely to be distorted, however, because of the applicants' perceptions that their answers could be verified easily.

Income was not used as a predictor (except for receipt or nonreceipt of public assistance) because it is precisely this variable that the participant is most likely to misreport. Income was, however, used in exploratory analysis to obtain a better understanding of the nature of misreporting. Correlations between the presence and absence of different sources of income and misreporting were calculated to determine whether any relationships existed. The analysis revealed that number of earners was highly correlated with misreporting. Number of earners was not used as a predictor variable, however, because this information was not requested on the 1981-82 application. As the next chapter indicates, this information can be helpful for making policy recommendations about the format of the application.

One additional consideration in the selection of predictors is face validity. Face validity is the intuitive judgment that a variable

should be indicative of misreporting. Some variables may correlate with misreporting through chance or through a complex network of causes and effects. Unless the variable can be credible as a predictor independent of its empirical properties, it would be difficult to implement the model. For example, grade of student may correlate significantly with misreporting by chance, because data were collected only from students in selected grades, or because there is a real relationship between grade level and misreporting. Because the variable does not have strong face validity, it might not be included in an error-prone model profile.

A complete list of all variables used and their results are contained in Appendix A. Those that appear in the EPM are operationally defined in Appendix C.

The Dependent Variable

The dependent variable used in error-prone model development is always some measure of error. Depending on the available data and the needs of the agency that will use the model, errors can be defined as discrepancies between actual and reported income, between correct eligibility and eligibility determined erroneously, and/or between benefit amounts awarded and the amount that should have been awarded based on accurate application data. For the school meal programs, error is defined as any instance in which a child appears to have received benefits in excess of those to which he/she is entitled, including:

- Receiving a free or reduced-price meal when he/she is not entitled to either benefit; or
- Receiving a free meal when he/she is entitled only to a reduced-price meal.

The determination of what the child received was based on the family's response in the NESNP survey, unless there were inconsistencies within the survey data. In cases with inconsistencies, the following rules were observed:

- If a respondent did not report receiving free lunch but reported receiving free breakfast, the target child in the family was classified as receiving free meals; and

- If a respondent reported receiving reduced-price lunches but reported paying more than the Federal ceiling (40 cents at the time) for the lunch, the target child in the family was excluded from the sample.

The formula used to approximate the free or reduced-price meal status to which a respondent was entitled involved comparing reported income to the eligibility scales. Adjustments were made to allow for situations in which households may have been given a hardship deduction[2] because of high housing expenses on other unusual expenses. If the applicant received more benefits than he/she was entitled to even after a hardship adjustment was taken into account, the application was classified as involving error.

Because the NESNP data base did not identify whether an applicant received a hardship deduction, the procedure used provided a generous interpretation of eligibility. In other words, a child determined to be eligible by the hardship formula may or may not have been eligible, but a child determined to be ineligible could not have been eligible under any interpretation of the regulations. A correction of this type will not be needed in the future because the enabling legislation no longer permits granting hardship deductions.

Development of the EPM

The statistical procedure used to generate the initial error-prone model was sequential search, using the Automatic Interaction Detector (AID) program developed by the Institute for Social Research at the University of Michigan. Appendix B discusses the specific parameters used in the analysis. AID has been used with considerable success by developers of EPMs for a number of agencies.

AID uses a hierarchical scheme that divides groups of cases in such a way that their differences are maximized. In simple terms, AID picks as its first predictor variable, the one that can break the total sample into two groups--one that has a large proportion of cases with errors, and a second that has a small proportion of cases with errors.

Each of the two groups is split again so that the difference between the proportions is maximized, and the process continues in an

iterative fashion until a subsample can no longer be split meaningfully because, for example, the number of cases in each of the two subsamples is very small or differences between the two subgroups are very small. Any subsample that cannot be split further becomes one of the groups defined by the model.

Validation of the Model

The results of the preliminary effort were validated using the replication sample. As noted above, the replication sample was a subset of the NESNP data base that was originally set aside to provide a separate data set for testing the predictive value of the preliminary groupings. Splits that were not validated by the replication sample were apparently due to chance. The variables that produced chance splits were dropped as a result of the findings from the replication sample and new splits that did not include these variables were created.

Results

The EPM development using NESNP data had three results: misreporting rates estimated from the NESNP data, variables associated with high error rates, and the model itself.

Misreporting Rates: Exhibit 1 presents findings on the extent of misreporting in the NESNP sample of households. It indicates that approximately 80 percent of those families reporting receipt of free meals (breakfast and lunch) were indeed, eligible for these benefits. The error rate was higher among those reporting receipt of reduced-price lunches; only 49.6 percent of this group was receiving the appropriatelevel of benefits. Because many more respondents reported receiving free meals than reduced-price meals, the total error rate was approximately 22 percent. The analysis reveals both cases of children receiving higher and lower benefits than those to which they are entitled; much of the error in the reduced-price category is associated with children paying for meals when they appear to be eligible for free meals.

EXHIBIT 1: ERROR RATES IN AWARD OF BENEFITS FOR SCHOOL NUTRITION PROGRAMS*/

CORRECT ELIGIBILITY STATUS	ELIGIBILITY STATUS AS DETERMINED BY SCHOOL OFFICIALS		
	Receiving Free Lunches	Receiving Reduced-Price Lunches	Receiving Free Breakfast
Receiving appropriate level of benefits	78.7%	49.6%	81.5%
Should pay full price	5.5	19.5	18.5
Should pay reduced price	15.8	NA	NA
Should receive free meals	NA	30.9	NA

*/Based on an analysis of the NESNP household survey of parents.

It should be noted that these findings are based on reports by parents about both their income and their child's eligibility status. Error rates were estimated without reference to the application or data on actual benefit status. Assumptions about the reasons for discrepancies between income level and eligibility status were made on the basis of decision rules noted earlier in this chapter. The findings suggest that personal interviews with parents, in which detailed income information was elicited, successfully obtained information about more income that was reported on the application.

The error rate found in this analysis is somewhat lower than the rate of error found in previous OIG studies. This may have occurred because reporting accuracy increased. A more likely explanation, however, is that the NESNP data base permitted a less restrictive estimate of error than did the OIG study. The OIG study used verified information. In this analysis errors were based on discrepancies in self-reported eligibility and eligibility based on self reported income and household data. However, the error rates estimated from NESNP are consistent with the finding of extensive misreporting in the school meal programs.

Variables Associated with High Error Rates: The analysis identified variables that appear to be associated with a high rate of errors in eligibility status and variables that seem to be associated with accuracy in eligibility status. Exhibit 2 presents a listing of the variables related to these outcomes. Because the results are based on a bivariate analysis that continually splits the sample into two groups (one with a higher proportion of errors and one with a lower proportion errors) the variables presented in the two categories are complements of each other.

Most of the variables that are related to a high error rate are indicators of being relatively well-off financially. For example, a higher proportion of error is found among households: (1) with income from dividends or interest; (2) with multiple wage earners; or (3) who own their own homes. These factors are indicators that family income is likely to be above the eligibility limits and also appear to be surrogates for carelessness or dishonesty in reporting income.

EXHIBIT 2: VARIABLES RELATED TO MISREPORTING

High Error Group	Error Rate %	Percent of Sample	Low Error Group	Error Rate %	Percent of Sample
Presence of dividends or or interest	51.9	2.7	No dividends or interest present	21.1	97.3
Two or more wage earners	41.8	34.0	One or no wage earners	12.9	65.0
Do not receive Food Stamps	35.5	53.6	Receive Food Stamps	6.2	46.4
Self-employed	34.1	4.1	Not self-employed	21.4	95.9
Presence of male adult in non-farm labor force	32.4	49.6	Male adult absent, farming or not in labor force	11.6	50.4
Male adult has some high school education	32.1	42.4	No male adult or male less than a high school education	14.4	57.6
Female adult absent or in non-farm labor force	32.2	46.5	Female adult farms or is not in labor force	12.4	53.4
No siblings in school or at least one receiving lower meal benefits	31.1	21.0	Some siblings in school and all receive at least some benefits	19.5	79.0
Own home	28.5	52.1	Do not own home	14.7	47.9
Do not receive public assistance	28.1	71.5	Receive public assistance	6.4	28.5
At most one sibling attends school	27.9	43.2	More than one sibling attends school	17.4	56.8
Male adult present	27.5	63.2	Male adult absent	12.3	36.8
Female adult is high school graduate	27.4	47.8	No female adult or female has less than a high school education	17.0	52.2

BEST COPY AVAILABLE

Indicators of lower income, such as receipt of public assistance payments or Food Stamps and the absence of a wage earner are associated with lower error rates. These measures are indicators that the household probably does have a relatively low income and is eligible to participate in school meal programs.

Exhibit 2 also shows the proportion of the sample that has each characteristic. For example, the likelihood of error in eligibility status is 51.9 percent for households with dividend or interest income but only 2.7 percent of the sample households have this type of income. When selecting variables for an error-prone model, both error rate and proportion of population must be considered because the most useful variables are those that are both characteristic of a high proportion of the population and associated with a high probability of error.

The most important findings are presented below.

- Household size, alone, is not a useful variable-- what is important is knowing whether the household contains a large number of children or adults. A larger number of children attending school tends to be associated with program eligibility and low misreporting whereas a larger number of adults in the household tends to be associated with higher income and higher misreporting.
- Source of income is an important variable because self-employment and employment in the nonfarm labor force tends to be associated with higher incomes and higher error rates whereas income from farming or transfer payments tends to be associated with a lower rate of misreporting.
- Food Stamp status is a valuable variable because receipt of Food Stamps tends to be associated with lower rates of misreporting.
- Variables that seem to be especially useful because they are associated with a high rate of error and are found in a large proportion of the population are: household not participating in the WIC program; presence of a male adult in the household; living in a single family dwelling; household not receiving public assistance; and household not receiving Food Stamps.

In summary, the analysis which prefaced the model development suggests that variables which are surrogates for income above poverty (e.g., home ownership, living in a single family dwelling, higher levels of educational attainment, income from wages, small family size, etc.)

tend to be associated with ineligibility and error. Program eligibility and accurate reporting tend to be related to surrogates for low income (e.g., participation in public assistance programs, large family size, little or no income from wages, single female head of household, unemployed male, etc.).

The Groups Generated by the Error-Prone Model: As was mentioned earlier, all of the variables that were analyzed were not considered in the development of the model. Instead, only those variables that met the following conditions were used:

1) They should be available on the application or easily accessible so that they could be used by school personnel in applying the model.

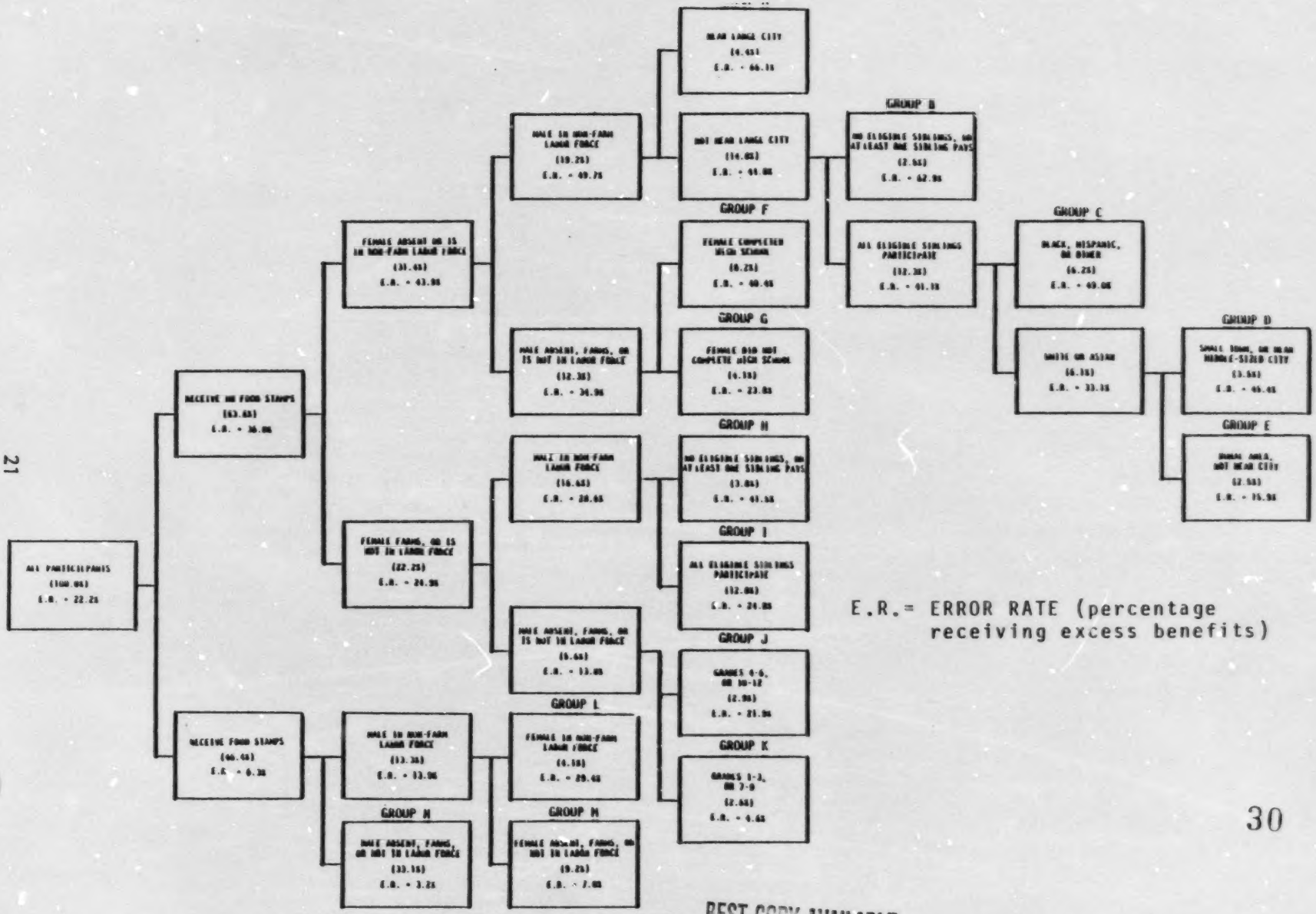
2) The accuracy of the variables available to school personnel should be the same as the accuracy achieved in NESNP. For instance, because income amount tends to be misreported on the application, and may not be misreported in the NESNP interview, income amount from NESNP would not be a useful variable for the EPM.

The initial error-prone model divided the sample into two groups--one with a higher proportion of errors and one with a lower proportion of errors. A total of 14 mutually exclusive groupings were generated in the initial model using variables that might be available to determining officials and were characteristic of a substantial proportion of the sample. These variables were selected because they would produce the most efficient scheme for locating misreporters.

Exhibit 3 is a tree-diagram that shows the subgroups that were produced by the initial error prone model. The statistic shown in parentheses is the percent of the sample possessing that combination of characteristics. The E.R. statistic is the error rate found among individuals in that grouping. For example, Group B is recipients who do not receive Food Stamps, where the female adult is absent or in the nonfarm labor force, the male adult is in the nonfarm labor force, the household is not near a large city, and there are no eligible siblings or the siblings pay for school meals. This group constitutes 2.5 percent of the sample and the error rate for the group is almost 70 percent.

A second summary of the groups is presented in tabular form in

EXHIBIT 3: ERROR-PRONE MODEL TREE DIAGRAM



BEST COPY AVAILABLE

Exhibits 4 and 5. Groups shown in Exhibit 4 have above average error rates; groups shown in Exhibit 5 have below average error rates. If the error-prone model were used to select suspicious applications, groups in Exhibit 4 would be selected for further verification of income data.

Conclusions about the Preliminary Error-Prone Model Developed from NESNP Data

This analysis produced important conclusions.

- First, it was possible to develop an error-prone model for the free and reduced-price school meal programs.
- Second, use of this error-prone model in selecting applications for verification would probably produce substantial savings over random selection of applications.
- Third, the analysis leading to the development of the model suggested several changes in the application and the application process that could contribute to a reduction in misreporting and misallocation of program resources.

It is important to note, however, that the results of this error-prone modeling development effort are very preliminary. School meal application requirements have changed since 1981 and new application procedures may elicit more accurate information. If verification requirements are implemented, they too, would alter misreporting patterns. These developments would probably outdate the specific model generated by this study. Therefore, the model developed in this analysis should not be implemented until it is tested with more recent IVPP data. However, the analysis performed in developing this model revealed important information about the nature of misreporting. It appears reasonable to use this information to improve the meal application process.

Evidence from the analysis suggests that changes to the application could reduce misreporting. First, the analysis strongly suggests that applicants must report detailed income information for each adult. This would be accomplished by providing separate lines for each adult household member to report income from each source (e.g., wages, pensions, welfare, etc.), and requiring that a response be placed on each line (including zero or none if no income is received from that source). Income may be underreported on the application because prior and current

EXHIBIT 4: GROUPS WITH ABOVE AVERAGE PROPORTION OF ERRORS

Group	Percent of Sample	Error Rate
Participants not receiving Food Stamps; female householder absent or in the non-farm labor force, male householder in non-farm labor force; near a large city.	4.4	66.1
Participants not receiving Food Stamps; female householder absent or in non-farm labor force; male householder in non-farm labor force; not near a large city; and no eligible siblings or at least one sibling pays.	2.5	62.9
Participants not receiving Food Stamps; female householder absent or in non-farm labor force; male householder in non-farm labor force; not near a large city; siblings attend school, but none pays for lunch; black, Hispanic or other.	6.2	49.0
Participants not receiving Food Stamps; female householder absent or in non-farm labor force; male householder in non-farm labor force; siblings attend school, but none pay for lunch; white or Asian; small town or near medium sized city.	3.5	45.4
Participants not receiving Food Stamps; female absent or in non-farm labor force; male absent, farms, or is not in labor force; female completed high school.	8.2	40.4
Participant not receiving Food Stamps; female absent or in the non-farm labor force; male absent, farms, or is not in the labor force; female did not complete high school.	4.1	23.8
Participant not receiving Food Stamps; female farms or is not in labor force; male in non-farm labor force; no siblings attend schools, or at least one sibling pays.	3.8	41.5

EXHIBIT 4: GROUPS WITH ABOVE AVERAGE PROPORTION OF ERRORS (Continued)

Group	Percent of Sample	Error Rate
Participants receive no Food Stamps; female farms or is not in labor force; male in non-farm labor force; some siblings in school and all participate.	12.8	24.8
Participants receive Food Stamps and both male and female householders are in non-farm labor force.	4.1	29.4

EXHIBIT 5: GROUPS WITH BELOW AVERAGE ERROR RATES

Group	Percent of Sample	Error Rate
Participants not receiving Food Stamps; female householder absent or is in non-farm labor force; male householder is in non-farm labor force; rural area not near city; some siblings in school but all pay; white or Asian.	2.5	15.9
Participants receive no Food Stamps; female farms or is not in labor force; male absent, farms, or is not in labor force; students in grades 4-6 or 10-12.	2.9	21.9
Participants do not receive Food Stamps; female farms or is not in labor force; student in grades 1-3 or 7-9.	2.6	4.6
Participants receive Food Stamps; male is in non-farm labor force; female is absent, farms or is not in labor force.	9.2	7.0
Participants receive Food Stamps; male householder is absent, farms, or is not in labor force.	33.1	3.2

(to a lesser degree) applications required the applicant to aggregate income for all adults. This method of requesting income information did not stimulate the respondent to consider all sources of income for all adults. In contrast, the NESNP income data and the in-home audits (discussed in the next chapter) queried respondents about each income source for each adult. The error-prone analysis findings, which showed the presence of multiple wage earners and multiple income sources to be indicative of misreporting, provide evidence to support this recommendation.

Two other changes could be helpful. One that is suggested by this analysis (and which was implemented by FNS in 1982-83) is requesting respondents to indicate whether they are receiving Food Stamp benefits. In addition to being useful for selective case actions, this question will also convey to applicants that a complete picture of the family resources is being obtained, which may help to deter misreporting. An additional general change to the application process is suggested by the error-prone analyses. If school districts require families to submit separate applications for each child, efforts should be made to cross-check the applications. A significant percent of families in the NESNP survey reported that children from the same household had different eligibility statuses for school meal benefits.

END NOTES

[1]This type of data was to be developed as a part of IVPP. EPM development using IVPP audit data is presented in Chapter 3 of this report.

[2]Although previously allowed, hardship deductions are no longer considered in eligibility determination.

3

REFINEMENT OF THE ERROR PRONE PROFILE BASED ON 1981-82 IVPP DATA

Although the analysis of the NESNP data did successfully produce an error-prone model, it had two major deficiencies. First, the best predictors of errors were based on a combination of variables, some of which are on the application and others which are not but are probably accessible. Use of the NESNP model would require time consuming screening on the part of school staff. A second deficiency in the NESNP model is that the definition of an erroneous application was based on self-reported data, not data verified in an audit. Data from the 1981-82 Income Verification Pilot Project were used to continue exploring the development of an error prone model. The IVPP-provided audited application data were used to answer the following questions:

1. Did the findings from the preliminary NESNP-based model replicate using the IVPP in-home audit data?
2. Could the NESNP model be refined, or a new model developed, that would be effective, yet very easy to use?

The remainder of this chapter presents the responses to these questions. The following topics are covered:

- The 1981-82 IVPP Sample,
- Statistical Methods and Results, and
- Conclusions.

The 1981-82 IVPP Sample

Because of the exploratory nature of the IVPP effort in 1981, the sample of school food authorities was not drawn randomly. Rather, an attempt was made to obtain variation among districts in terms of size, location and urbanicity. Consequently, the statistical generalizability

of findings from 1981-82 IVPP may be questioned. However, key findings from previous analyses[1] tended to be consistent across SFAs, suggesting a lack of strong bias and implying that the findings are probably somewhat generalizable.

The error-prone model analysis used a random sample of 741 families from 54 schools in nine SFAs that participated in Phase I of IVPP. These 741 families participated in personal interviews during which they were asked to provide documents (e.g., pay stubs, letters from employers, etc.) that indicated the amount of income received from 17 sources (e.g., wages, pensions, alimony, etc.) during the month in which they applied for benefits. In addition, household size was reviewed. These in-home audits, therefore, produced verified information from which eligibility could be computed and compared with the eligibility based on the information reported on the application for meal benefits. Cases of excess benefits occurred when in-home audit data yielded a lower eligibility rating than did the application data.

Statistical Methods and Results - Verification of NESNP Model

The analysis of the IVPP in-home audit data first addressed the validation of the NESNP findings. Each respondent in the in-home audit sample was assigned a score (log of the odds ratio) reflecting the likelihood of receipt of excess benefits as predicted by the NESNP model. The score was then correlated with whether or not the applicant had actually received excess benefits. This resulted in an index of discriminant power. The correlation was 0.46 for the original NESNP profile. In comparison, when application month data from the in-home audit were used, a correlation of 0.36 was obtained. A second correlation was produced using in-home audit month data. This produced a correlation of 0.48. The second correlation is very close to that from the NESNP data. Therefore, the profile as a whole replicates findings from the in-home audit data base. This important point should be noted: the correlations produced with in-home audit month income and family size data are much stronger than those produced with application month data. This result probably occurred because both the NESNP survey data and the in-home audit month income data were collected at the same time--during

the spring--about nine months after the applications had been submitted to school officials for approval. Therefore, the NESNP error-prone profile reflects a compounding of the effects of initial misreporting on meal benefit applications and eligibility status change during the course of the school year.

Examination of error rates for the various error-prone groups (that is, those groups with a strong likelihood of receiving excess benefits) must be approached cautiously because the in-home audit sample size is too small to properly verify the smaller splits (groups) in the model. However, the primary splits can be verified. Exhibit 6 presents the results of the validation of the groups. Each bar represents the proportion of each group receiving excess benefits. The two primary splits were receipt of Food Stamps and female adult household member in the labor force. The validation analysis of the NESNP error-prone profile has an important implication for interpreting findings from the in-home audits. The unrepresentative sample of SFAs and schools from which the in-home audit sample was drawn raised questions as to the generalizability of findings from the sample. The NESNP sample, by contrast, is nationally representative and was constructed in accord with generally accepted statistical principles. Therefore, the fact that the two surveys replicate on principal correlates of receipt of excess meal program benefits suggests that other findings from the in-home audits may also replicate nationally.

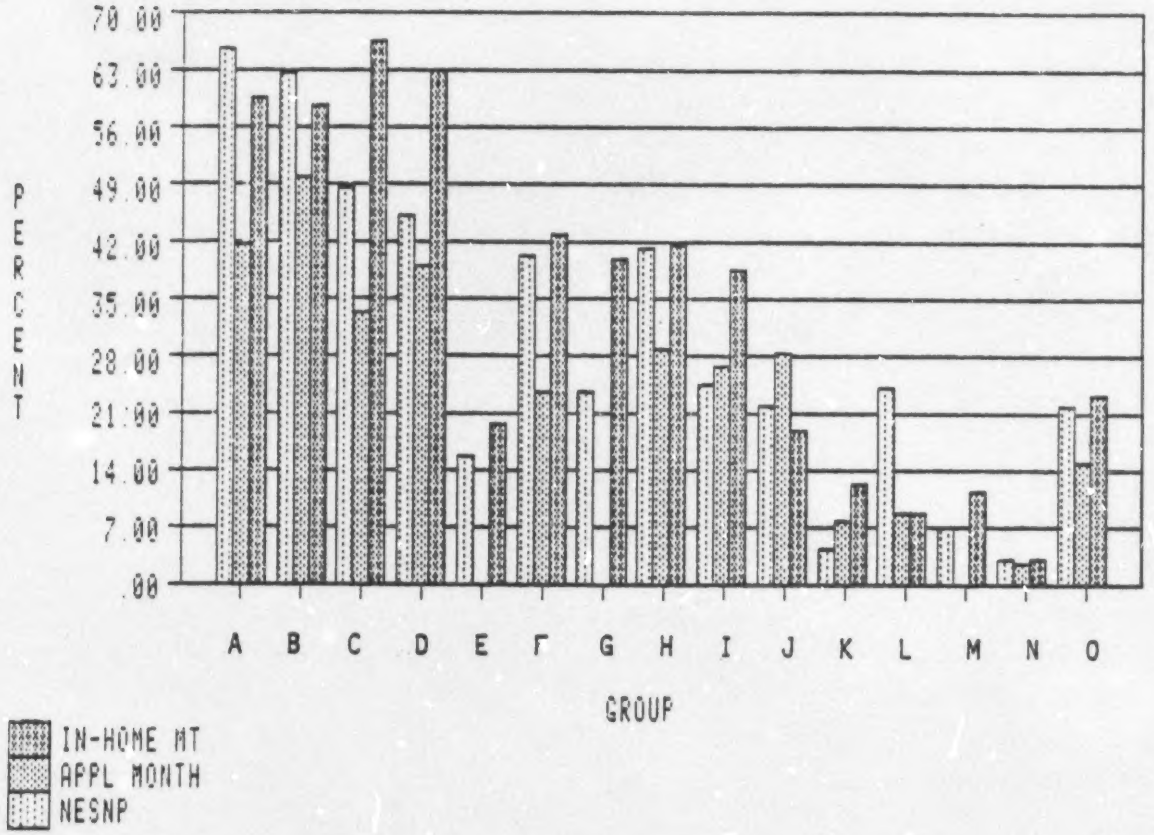
Statistical Methods and Results - An Improved Model Based on IVPP Data

The next issue addressed by the analysis was whether or not an EPM could be produced that would not be very burdensome for an SFA to use. It was decided that a model would be reasonable for implementation if it:

- relied on the most accessible information; that is, the information reported on the application; and
- involved very few procedural or screening steps.

Therefore, this effort excluded many of the variables (e.g., nonfarm labor force, sibling's participation status) that were effective predictors in the NESNP model EPM because they were not currently

EXHIBIT 6: VALIDATION OF NESNP-DEFINED ERROR-PRONE GROUPS USING NESNP DATA AND IN-HOME AUDIT DATA



KEY: A. Participants not receiving food stamps; female householder absent or in the non-farm labor force; male householder in non-farm labor force; near a large city.

B. Participants not receiving food stamps; female householder absent or in non-farm labor force; male householder in non-farm labor force; not near a large city; and no eligible siblings or at least one sibling pays.

C. Participants not receiving food stamps; female householder absent or in non-farm labor force; male householder in non-farm labor force; not near a large city; siblings attend school, but none pays for lunch; black, Hispanic, or other.

D. Participants not receiving food stamps; female householder absent or in non-farm labor force; male householder in non-farm labor force; siblings attend school, but none pays for lunch; white or Asian; small town or near medium-sized city.

E. Participants not receiving food stamps; female householder absent or in non-farm labor force; male householder in non-farm labor force; rural area not near city; some siblings in school but all pay; white or Asian.

F. Participants not receiving food stamps; female absent or in non-farm labor force; male absent, farms, or is not in labor force; female completed high school.

G. Participants not receiving food stamps; female absent or in the non-farm labor force; male absent, farms, or is not in the labor force; female did not complete high school.

H. Participants not receiving food stamps; female farms or is not in labor force; male in non-farm labor force; no siblings attend school, or at least one sibling pays.

I. Participants receive no food stamps; female farms or is not in labor force; male in non-farm labor force; some siblings in school and all participate.

J. Participants receive no food stamps; female farms or is not in labor force; male absent, farms, or is not in labor force; students in grades 4-6 or 10-12.

K. Participants do not receive food stamps; female farms or is not in labor force; male absent; student in grades 1-3 or 7-9.

L. Participants receive food stamps and both male and female householders are in non-farm labor force.

M. Participants receive food stamps; the male is in the non-farm labor force; the female is absent, farms, or is not in the labor force.

N. Participants receive food stamps; the male householder is absent, farms, or is not in the labor force.

O. Total

BEST COPY AVAILABLE

requested or anticipated to be requested on the application. The variables that are available on the application and were tested as potential predictors are shown in Exhibit 7.

The only application predictor of receipt of excess benefits that replicated in all nine sampled SFAs was what we call "threshold reporting." "Threshold reporting" is reporting a monthly income that is on or near the free or reduced-price eligibility cut-off point. For example, a family of five is eligible for reduced-price benefits if it has a monthly income of $1,516 or less. If a family of five reported its monthly income to be $1,506, it would, under our definition, be a "threshold reporter." Exhibit 8 displays the relative frequency of receipt of excess benefits by difference between eligibility cut-off point for benefits received and reported income. Seventy percent of those who reported incomes within $60 a month of the eligibility line receive excess benefits compared with only 10 percent of those who reported incomes more than $180 a month from the eligibility cut-off. (The $60 intervals used are not arbitrary but were constructed to maximize the discriminant power and robustness of differences.)

The 1981-82 application form used in this study did not contain information on receipt of Food Stamps. However, the FNS school meal benefit application recommended for the 1982-83 school year does request this information. Therefore, it was appropriate to examine receipt of Food Stamps relative to receipt of excess benefits, even though Food Stamp participation could only be identified during the in-home audit. Food Stamp participation data can be used to improve significantly the prediction of receipt of excess benefit reporting. Use of Food Stamp participation data in the error-prone model rests on the assumption that such information will not be significantly misreported on meal benefit applications. This assumption may be reasonable on the grounds that applicants for school meal benefits have no overt motivation to misreport food stamp participation. Eligibility for school meals is based solely on income and family size. It is possible, however, that some applicants incorrectly believe that Food Stamp participation does affect eligibility and therefore, may not report it on the application.

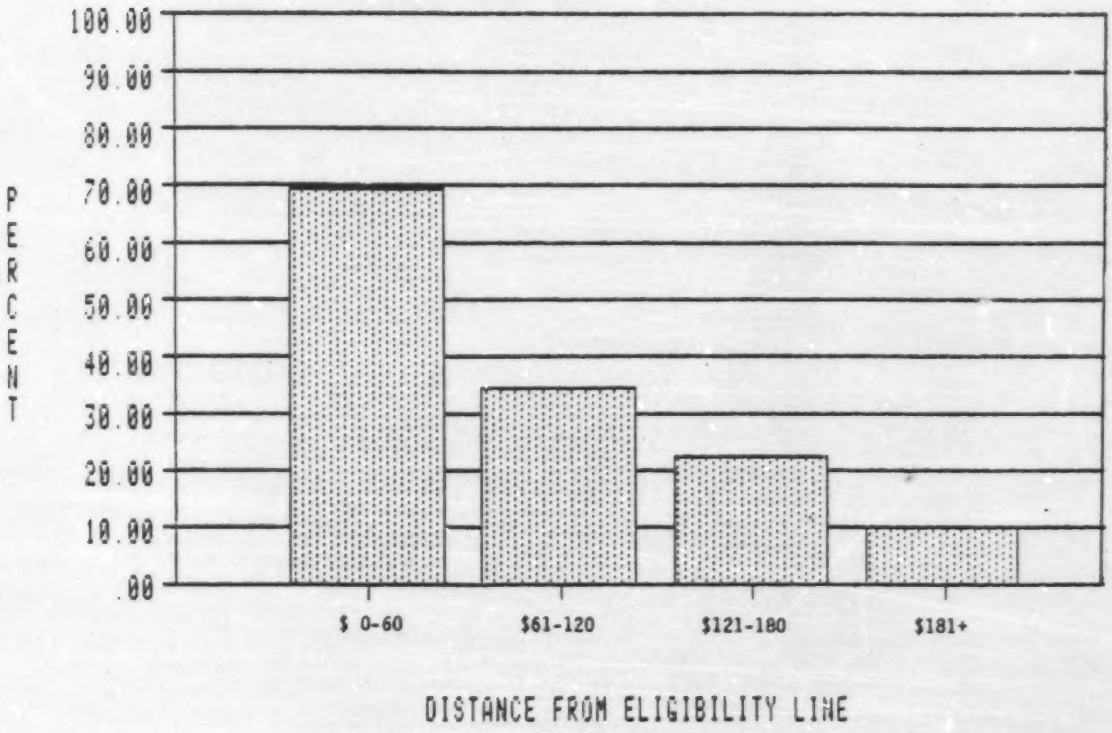
EXHIBIT 7: LIST OF VARIABLES USED IN DEVELOPMENT OF APPLICATION-BASED ERROR-PRONE PROFILE

Income from wages and other earnings
Income from Social Security
Income from public assistance
Income from unemployment benefits
Income from child support or alimony
Income from pension or retirement
Other income
Family size
Total reported income
Eligibility status based on application
Difference between reported income and eligibility cut-off point
Whether income was reported in round numbers (evenly divisible by 25, 50, 100)
Presence, or absence of Social Security numbers
Grade of student
Receipt of Food Stamps*
Eligibility based on in-home audit*
Difference between application and in-home audit based eligibility*

---

*Data from in-home audit used. All other variables from application.

EXHIBIT 8: PERCENTAGE OF HOUSEHOLDS RECEIVING EXCESS BENEFITS BY DIFFERENCE OF REPORTED MONTHLY INCOME FROM ELIGIBILITY THRESHOLD



Food Stamp participation status was added to the application variables shown in Exhibit 7 and a new error-prone model was developed through a six-stage process. First, bivariate relationships between excess benefits and application variables were examined and the application variables were rescaled to maximize predictive power. Second, a step-wise logistic discriminant function model was estimated to select the most predictive set of variables. Third, a sequential search using automatic interaction detection (AID) was conducted to assure that no hidden interactions existed among the variables that were not discovered by previous steps. Fourth, a prediction model was estimated using a weighted least-squares procedure that optimized discriminant power. Fifth, the prediction equation was simplified to allow easy use by SFA officials. Sixth, the prediction equation was tested separately for each of the nine sampled SFAs.

The resulting error-prone model is very simple and contains only two variables. Exhibit 9 presents the scoring profile derived from the model. An application is given one point if it reports an income within $120 a month of the free or reduced-price eligibility cut-off. Another point is added if the reported income is within $60 of the free or reduced-price eligibility cut-off. Finally, a point is subtracted if the applicant reports receiving Food Stamps. The resulting scale has values ranging from -1 to +2. Exhibit 10 displays the percentage of all sampled applicants receiving excess benefits by error-prone score and the percentage of the sample in each of the score categories. In each of the nine sampled SFAs, the error-prone model significantly discriminated between applicants receiving excess benefits and all other applicants.

In addition to being easy to use, this model has the following beneficial features:

- It is able to separate nearly half the applicants into a group of accurate reporters (45% of the sample received Food Stamps and only 2% of these Food Stamp recipients received excess meal benefits). Identification of accurate reporters is helpful because it prevents the waste of pursuing verification of these applications.
- The 6 percent of applicants who receive an error-prone score of "2" have applications that are four times as likely to contain an error that would reduce eligibility status as an

EXHIBIT 9: APPLICATION-BASED ERROR-PRONE PROFILE SCORING SYSTEM

Step		
A.	If reported income is within $120 a month of the free or reduced-price eligibility cut-off, write '1' on line A. Otherwise write '0.'	_____(A)
B.	If reported income is within $60 a month of the free or reduced-price eligibility cut-off, write '1' on line B. Otherwise write '0.'	_____(B)
C.	If the applicant reports receiving Food Stamps, write '-1' on line C. Otherwise write '0.'	_____(C)
D.	Sum lines A, B, and C, and write final score on line D.	_____(D)

EXHIBIT 10: APPLICATION-BASED ERROR-PRONE PROFILE

Error-Prone Score*	Percentage Receiving Excess Benefits	Percentage of Total Sample
-1	2%	45%
0	20%	42%
1	40%	7%
2	71%	6%

* Error prone scores were derived statistically by a weighted least squares procedure. A linear transformation was performed to preserve the relative weights, but produce whole numbers.

applicant selected randomly from the total sample.

The final step in developing the application-based error-prone model was to analyze what types of error are, and are not, detectable by the profile. The error-prone model relies heavily on threshold income reporting. Therefore a particular concern in testing the model was to determine whether the model detects only small reporting errors made by households with actual incomes near their reported incomes. If this were the case, then the error-prone model would fail to detect large errors made by more affluent households.

To address this issue, misreporters identified by the scoring profile were compared with misreporters not identified by the profile in terms of how much greater their actual income was than the income cut-off for the level of program benefits they received. The purpose of this comparison was to determine whether the error-prone profile is biased against low-income households whose actual monthly incomes are close to the eligibility cut-off. Detectable misreporters had median monthly incomes $230 above the maximum allowable income for benefits received. This compares with a median monthly income of $255 a month above the maximum allowable income received by nondetectable misreporters. Thirty-four percent of detectable misreporters had actual monthly incomes within $100 of the maximum allowable for benefits received compared with 30 percent of nondetectable misreporters. The difference of 4 percent was not statistically significant. Therefore, the model does not appear to discriminate againsst those who are truly, but marginally, eligible.

Summary and Conclusions

The 1981-82 IVPP in-home audit sample was used to refine the preliminary model developed from NESNP data. The results showed that the model developed from the NESNP data base was validated on the in-home audit data. This result produces two important conclusions:

- Results from a comprehensive review of income by source in a personal interview that occurred in NESNP produced results similar to results from the in-home audit which required documentation to verify income. This finding strongly suggests that more accurate income information would be obtained if the free and reduced-price meal application requested income by

source, to stimulate the applicant to consider all sources of income for each adult.

- The NESNP sample was statistically representative and the IVPP was not. The replication of the NESNP findings on the IVPP data suggest that the IVPP results are probably generalizable beyond the sample of SFAs that participated.

The goal of producing a simple but effective EPM also was achieved. The scoring profile, which is easy for SFA officials to use, is capable of selecting a subgroup of applicants for verification that have a 70 percent probability of receiving excess benefits. The error-prone profile was successful in selecting applicants receiving excess benefits in all nine sample SFAs. Use of the error-prone profile for a six percent or smaller sample would, on the average, result in selection of applications for verification that have up to four times the probability of receiving excesss benefits as applications selected at random.

In 1982-83, IVPP is testing more sophisticated verification procedures in a nationally representative sample of SFAs, and will include a large in-home audit sample. Application improvements implemented in 1982-83 may outdate the model, and the sophisticated verification techniques may alter misreporting patterns. Thus, it will be necessary to further examine the EPM developed in 1981-82 with the 1982-83 IVPP data.

END NOTES

[1]Income Verification Pilot Project, School Year 1981-82 In-Home Audit Findings, Applied Management Sciences, May 1983; and Findings on School Meal Program Participation, Applied Management Sciences, December 1982.

APPENDIX A

PREDICTORS USED IN THE ANALYSES

Predictors	Code[1]
1. Local density (city, suburb, small town, rural)	1
2. Global density (near large city, near middle-sized city, small town, not near city)	1
3. Number of children attending school	1
4. Number of children not attending school	2
5. Type of dwelling	2
6. Location of dwelling (acres surrounding home)	2
7. Ethnicity of target child	1
8. Owns home	1
9. Highest grade attained by male householder	1
10. Highest grade attained by female householder	1
11. Work status of male householder (farm, non-farm, absent, or not in labor force)	1
12. Work status of female householder (farm, non-farm, absent, or not in labor force)	1
13. Number of adults in household	2
14. Grade of child (1-3, 4-6, 7-9, 10-12)	1
15. Participation of eligible siblings (no eligible sibling, all participate, one pays)	1
16. Food Stamp participation	1
17. Receives food under WIC	2
18. Receives welfare payments	2
19. Relationship of target child to mother	3
20. Relationship of target child to father	3
21. Highest grade attained by either householder	3
22. Participation in previous year	3
23. Number of children 1-5	3
24. Number of children 6-13	3
25. Number of children 14-18	3
26. Number of unrelated adults	3
27. Number of related adults	3
28. Income by wage-earner	3
29. Self-employed, non-farm	3
30. Self-employed, farm	3
31. Income from dividends or interest	3
32. Social Security income	3
33. Number of older siblings	3
34. Number of younger siblings	3
35. Age of oldest sibling	3
36. Age of youngest sibling	3
37. Age of child	3
38. Age of father	3
39. Age of mother	3
40. Occupation of father	3
41. Occupation of mother	3
42. Sex of child	3

[1]Codes are listed below

1 = Appears in model
2 = Used in AID run
3 = Examined bivariately

APPENDIX B

ANALYTIC PARAMETERS

Analytic Parameters of the AID run were:

-- One dichotomous criterion variable.

-- Eighteen predictors--seven free and 11 monotonic--for the Food Stamp Model.

-- Reducibility criterion = 0.002.

-- Minimum group size=40.

To retain a split, the phi coefficient (equivalent to the Pearson r) between the dichotomous split and the criterion variable in the cross-validation sample had to be large enough to have been significant for the working sample, and of similar directonality.

APPENDIX C

OPERATIONAL DEFINITIONS

The following define operationally the splits produced by the AID analysis for the two models. For clarification of the predictor variables, NESNP documentation associated with the data collection should be consulted.

1. Male is in non-farm labor force: HQ128 is present and not equal to 9, 10, or 13.

2. Male is absent, farms, or is not in labor force: HQ128 is missing or equal to 9, 10, 13.

3. Female is absent or is in non-farm labor force: HQ136 is missing or is present and not equal to 9, 10, or 13.

4. Female farms or is not in labor force: HQ136 is equal to 9, 10, or 13.

5. Female is in non-farm labor force: HQ136 is present and not equal to 9, 10, or 13.

6. Female is absent, farms, or is not in labor force: HQ136 is missing or equal to 9, 10, or 13.

7. Near large or middle-sized city: HQ184 is 7 or less.

8. Small town or rural area, not near city: HQ184 is 8 or 9.

9. No eligible sibling or at least one sibling pays: None of HQ121-4 to HQ121-12 is 1, or for one of the above mentioned variables the value is 1 and the value of the corresponding HQ123 variable is greater than 0 (or greater than 40 if reduced-price participant).

10. All eligible siblings participate: At least one of HQ119-4 to HQ119-12 is 1 and whenever this happens, the corresponding HQ123 variable is 0 (or not greater than 40 if reduced-price participant).

11. Black or Hispanic: HQ180 is 2 or 3.

12. White or Other: HQ180 is 1 or 5.

13. One or no siblings attend school: Number of HQ119-4 to HQ119-12 equal to 1 is not greater than one.

14. Two or more siblings attend school: Number of HQ119-4 to HQ119-12 equal to 1 is greater than one.

15. Small town: HQ184 is 8.

16. Rural area, not near city: HQ184 is 9.

17. Male has not completed high school: HQ127 less than or equal to 2 or missing.

18. Male has completed high school: HQ127 greater than 2.

19. Cities, suburbs, or towns: HQ184 is 1, 2, 3, 5, 6 or 8.

20. Rural areas: HQ184 is 4, 7, or 9.

21. Owns house: HQ164 is 1.

22. Does not own house: HQ164 is not 1.

23. Receive no Food Stamps: HQ147 is not 1 and HQ151 is not 1.

24. Receive Food Stamps: HQ147 is 1 or HQ151 is 1.

25. Near large city: HQ184 is 1, 2, 3, or 4.

26. Not near large city: HQ184 is 5, 6, 7, 8, or 9.

27. Black, Hispanic, or other: HQ180 is 2, 3, or 5.

28. White or Asian: HQ180 is 1 or 4.

29. Small town or near middle-sized city: HQ184 is 5, 6, 7, or 8.

30. Female completed high school: HQ135 is greater than 2.

31. Female did not complete high school: HQ135 is 1 or 2 or missing.

32. Grades 4-6 or 10-12: GRADE is 4, 5, 6, 10, 11, or 12.

33. Grades 1-3 or 7-9: GRADE is 1, 2, 3, 7, 8, or 9.